

DEVELOPING AND ASSESSING IMMERSIVE CONTENT FOR NAVAL TRAINING: LESSONS LEARNED IN THE VIRTUAL WORLD

Jason H. Wong, Ph.D., Naval Undersea Warfare Center, Newport, RI
Anh B. Nguyen, Naval Undersea Warfare Center, Newport, RI

Content was developed in an immersive, 3-D virtual world (SecondLife Enterprise) as a pilot project to support a Navy training need to teach complex, highly visual concepts in an engaging fashion. An avatar-based virtual world was used due to their unique social and content creation properties. This project consisted of a series of modules designed to teach the naval concept of target motion analysis. The resulting “immersive learning trail” augmented a two-week unit that was part of a longer course. Students from a Naval schoolhouse participated in classes that used this content, and several assessments were performed. Students gained the most from the content when complex spatial concepts were taught, though impressions of the entire trail were positive. Several lessons learned are reported, including the need to consistently make the experience as immersive as possible and the kind of content that benefits most from immersion.

INTRODUCTION

The explosion of virtual environment technology has enabled the creation of content that simulates and immerses the user in a digital representation of a real-world environment. These platforms have been leveraged by the military training community to create scenario simulation and serious games that aim to enhance training effectiveness.

A typical example of the work done in serious games is the Navy Flooding Control Trainer. This system was developed in 2009 in order to supplement classroom instruction and to reduce the demand on training resources at the Navy Recruit Training Command. This trainer was developed with an open source game engine, and the focus was to provide a 3-D environment of a simulated Arleigh-Burke class destroyer for individual flooding control training.

Usability and validation studies were conducted and showed that all participants in the virtual training group navigated to the correct deck in a flooding situation. Only 67% of control group participants followed the correct procedure. Similarly, 36% of the treatment group described the leak correctly (versus 16% of the control group), and only 28% of the treatment group entered the flooding compartment without appropriate personal protective equipments (compared to the 67% in the control group). These results demonstrated that the FCT provided sailors with the ability to acquire a strong ship damage control foundation above and beyond the traditional training material (Murphy & Hussain, 2009).

While serious games and scenario simulations comprise a majority of virtual trainers (Murphy, 2011), avatar-based virtual worlds provide several unique advantages to content developers, instructors, and students without wholly focusing on scenario simulation. Virtual worlds provides a social platform for users and the ability for anyone to create content that persists across time. They are also suited to augment initial learning by providing concept exploration, while scenario simulations are typically focused on providing students with procedure practice. Virtual world-based training was hypothesized to have maximum impact on learning effectiveness when the content represented complex, spatial information and augmented other curriculum components.

In order to test these ideas, the Naval Undersea Warfare Center partnered with a U.S. Navy schoolhouse to develop

content designed to augment already-existing course material designed to teach the target motion analysis (TMA), which is a process used to estimate the course, speed and range of a target in the water using sonar data. This material was part of a longer course for novice submarine fire control technicians (FTs), whose primary role is contact management. This content aims to provide an understanding of the fundamentals of the TMA problem and link that knowledge to current submarine combat systems. This will be done by helping students comprehend how ground truth data is gathered, processed, and displayed by various combat system displays in order to solve the TMA problem.

An iterative development process was implemented, with feedback from Naval instructors continually integrated into the content. An assessment plan using several novice FT classes was developed and executed, and multiple kinds of data were collected, from surveys to test scores to a sense of immersion. Results show that the modules that best enhanced learning dealt with the most complex, difficult-to-visualize concepts.

CONTENT DEVELOPMENT

Initial discussions occurred between the content developers and the Naval schoolhouse seeking to identify overlap between trouble spots in the curriculum and available developer expertise on the subject matter. The Solution Sensitivity (SS), used by FTs to perform TMA tasks, was identified as a difficult-to-understand concept that contained complex spatial elements ready for visualization.

The first stages of development were devoted to creating an immersive SS plot. TMA experts provided input and feedback, and various iterations of the exhibit were created. A final optimized version was completed with the appropriate scale based on the average size of an avatar. This large scale was chosen because one design goal was to immerse avatars in the SS plot, and this scale allowed avatars to walk into the plot and interact with the individual elements.

Additional feedback from the Naval schoolhouse and curriculum developers were positive, and they asked for additional exhibits to be created in order to provide the necessary background to help students understand the context behind SS. The content of the remaining exhibits was built

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE OCT 2012		2. REPORT TYPE		3. DATES COVERED 00-00-2012 to 00-00-2012	
4. TITLE AND SUBTITLE Developing And Assessing Immersive Content For Naval Training: Lessons Learned In The Virtual World				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Undersea Warfare Center,Newport,RI,02841				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES To be presented at the HUMAN FACTORS and ERGONOMICS SOCIETY 56th ANNUAL MEETING - October 22-26, 2012, Boston, MA					
14. ABSTRACT Content was developed in an immersive, 3-D virtual world (SecondLife Enterprise) as a pilot project to support a Navy training need to teach complex, highly visual concepts in an engaging fashion. An avatar-based virtual world was used due to their unique social and content creation properties. This project consisted of a series of modules designed to teach the naval concept of target motion analysis. The resulting ?immersive learning trail? augmented a two-week unit that was part of a longer course. Students from a Naval schoolhouse participated in classes that used this content, and several assessments were performed. Students gained the most from the content when complex spatial concepts were taught, though impressions of the entire trail were positive. Several lessons learned are reported, including the need to consistently make the experience as immersive as possible and the kind of content that benefits most from immersion.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 6	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

based on feedback from SMEs agreeing on the fundamental knowledge was necessary to understand SS.

A review of current course materials related to the TMA problem was undertaken, which included PowerPoint presentations and exams from various instructors. This provided the developers a baseline as to what the students were learning so that they could ensure the content would augment the currently-existing material. After the initial content was created, an iterative design process was used to provide continual feedback between content development team, curriculum developers and course instructors. After a three-month process, all parties were satisfied with the content. The resulting content consisted of four exhibits. Concerns about classification issues prevents the inclusion of screenshots from the actual exhibits and precise description of the concepts they were designed to teach. Approximations are shown to present the gist of the content.

The first exhibit deals with lines of sight (LOS). For readers unfamiliar with drawing bearing lines, Figure 1 depicts a scenario with two ships moving and lines of bearing drawn between them at various points in time. This exhibit was designed to teach the positional relationships between one's own ship (ownship) and a contact. The actual exhibit was similar to Figure 1 and contained an interactive controller that allows the students to view the scenario at their own pace. As the scenario progresses, lines of bearings are generated and drawn right on the water.

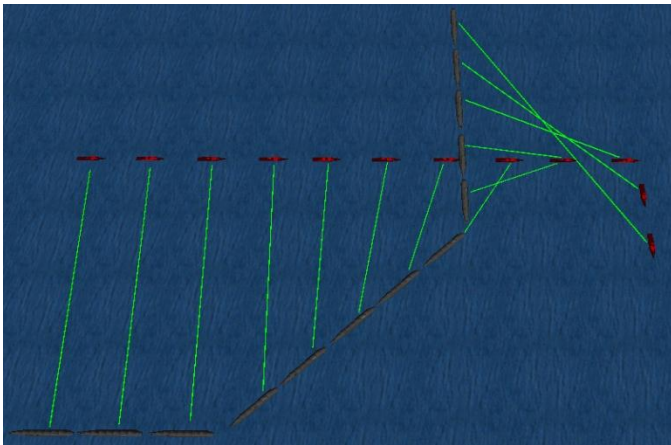


Figure 1. Example lines of bearing between two ships.

The second exhibit (not pictured) demonstrates Method 1. The goal of this analysis method is to try and fit the observed lines of sight with predefined patterns of course and speed. Before automated algorithms to accomplish this task were developed, operators had to manually attempt to fit patterns on these recorded lines of bearing, and errors had to be judged by the operator. The immersive exhibit has digital representations of the paper patterns where students could try and manually find the best fitting solution by trying one pattern at a time.

The third exhibit (not pictured) introduces Method 2. This is a visualization of the errors of fit produced by a single pattern solution that is trying to fit the data. This visualization is a result of algorithms used by the combat system and is useful because the pattern of errors in how well the solution fits the data is informative to the FT. The interactive portion of

this exhibit allows students to try out different solutions and observe how this information is correlated between the collected data and resulting errors.

Figure 2 shows an older version of the SS exhibit, which is designed to introduce the SS plot and link it to previous displays. The goal of SS is to automate the analysis of hundreds of different solution patterns simultaneously and then display the errors of those solutions in an intuitive fashion. On the combat system, the amount of error is represented through different colors. The immersive content takes advantage of the 3-D nature of the virtual world by giving height to the color-coded squares. Now, the height of each bar represents the amount of error for that particular solution. The interactive portion of this exhibit comes from students being able to examine how the SS plot changes over time due to the collection of additional data.

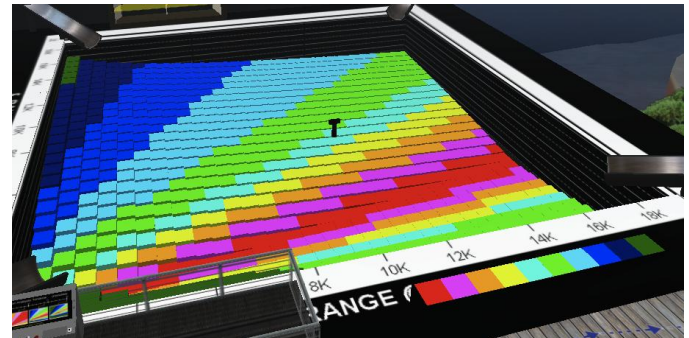


Figure 2. Immersive SS exhibit.

The entire trail consisted of these four exhibits. To summarize, the goal of the exhibit was to help students understand how data about ground truth was collected and analyzed into a colorful, complex display. The first exhibit introduced the data being collected, the second exhibit introduced the idea of fitting patterns to data, the third exhibit showed how errors of fit were displayed, and the final exhibit demonstrated how the system could simultaneously analyze many solutions and display error. If students could understand these concepts, then the objectives of the content were met.

CONTENT ASSESSMENT

Assessing the learning effectiveness of this content required a scientifically valid methodology while working within the constraints of the Naval schoolhouse. The current 18-week course material could not be interfered with, so time in the TMA trail was assigned as mandatory self-study. Additionally, all students within one class had to receive the same material, so one class could not be split into experimental and control groups – different classes of students had to be assigned to these groups.

Participants. As one class was designated to receive the experimental content, one instructor was recruited to help teach the content. This instructor had two years of teaching experience and was given about four total hours of training in the virtual world and additional content familiarization.

All student participants were novice FTs in the same 18-week class. Students were all male, aged 19-24. A total of four classes were used for this assessment. The experimental class

(16 students) was assigned mandatory self-study in the immersive TMA trail, which they used for a total of four hours across two sessions. One control class ("Control-Self Study", 11 students) was assigned four hours of mandatory self-study, where they had access to their traditional class materials and the instructor. Finally, two control classes ("Control-No Self Study/NSS", 12 and 9 students) did not receive any extraordinary mandatory study time. Two groups were needed due to a timing error where the pre-test given to students was given at a different point in the course compared to the other classes, so those results were unusable. The necessary test, survey, and other data was collected between these two classes to satisfy the Control-NSS condition.

Lab Setup. The study hall consisted of one SecondLife Enterprise server from Linden Labs, which served to host the content on a secure, private network. Students and the instructor logged into this server to access the TMA trail. There was one instructor workstation that projected the world onto a screen in the front of the classroom. Additionally, there were eight student computers facing the front of the classroom, each with a minimum level of specifications to smoothly run SecondLife at a medium level of graphics quality. Students had to pair up to use one workstation, but sharing the duties of avatar control was encouraged.

Procedure. The control groups all received minimal interference. The self-study group was assigned four hours outside of regular class time to study their own materials. All control groups also received: (1) a pre-test just after the start of the traditional TMA material, (2) a post-test before students were brought into the combat system trainers, and (3) a retention test at the end of the 18-week class.

The experimental group received these three tests at the same time during their 18-week class, and they also went through the TMA trail across two sessions during the two-week TMA module in the standard curriculum. During the first session, students were led through a SecondLife introductory trail, where they learned how to navigate in the virtual world, control the camera view, and teleport between points of interest. After this trail was completed, students teleported to the main trail, where they were led through the line of sight and Method 1 exhibits by the instructor. During the second session, students worked through the remaining Method 2 and SS plot exhibits. Afterwards, students took a post-test, likeability survey, and presence questionnaire. This entire session took approximately two hours. Finally, near the end of the 18-week class, they took a retention test.

During each of the class sessions, one of the content developers controlled the instructor's avatar in the virtual world, allowing the instructor to freely discuss the content and provide instructions to the "driver" as to where to move the avatar to next or focus the camera on a certain view. This was done because the instructor did not feel fully comfortable in the virtual world. During internal pilot testing, it was determined that someone with more experience in world could easily maneuver the avatar and lecture simultaneously.

One novel feature built by the developers was the ability of the instructor to force the camera view of the students to mirror that of the instructor. Typically, any avatar can control their camera as they wish, but some students had trouble

controlling the camera, and a common focus for all students was often necessary. Therefore, the instructor could force students' cameras to mirror his view. After the instructor was finished lecturing, students were encouraged to interact with various elements of each exhibit while the instructor walked around and asked questions.

Materials. Several pieces of data were collected from students to perform an assessment of learning (Kirkpatrick, 1998). The likeability survey was a Likert scale asking students about their impression of the usefulness of the entire trail and individual exhibits. The presence questionnaire was based on Witmer & Singer's (1998) scale of different factors of presence. The pre, post, and retention tests were developed by the researchers and Naval schoolhouse instructors and included questions asking about concepts from the four exhibits. Scores from standardized exams developed by the Naval schoolhouse were collected, along with scores from the student's Armed Services Vocational Aptitude Battery test (i.e., the military entrance exam).

RESULTS

Likeability/Usability Survey. Students generally found all of the exhibits useful (Figure 3) it must be noted that it is possible that these scores are a combination of the usefulness of the content itself and the accompanying instructor lectures, since this cannot be used without an instructor. Particularly high scores were found for the Method 2 and SS exhibits. This is especially positive because these are concepts that are generally considered difficult and relevant to today's Navy.

Students were additionally asked about how easy it was to move in the virtual world and how well they were able to find their way. Scores for both of these usability metrics was relatively high ($M = 4.75$, $SD = 1.53$ for ease of movement, $M = 5.13$, $SD = 1.36$ for ease of finding your way).

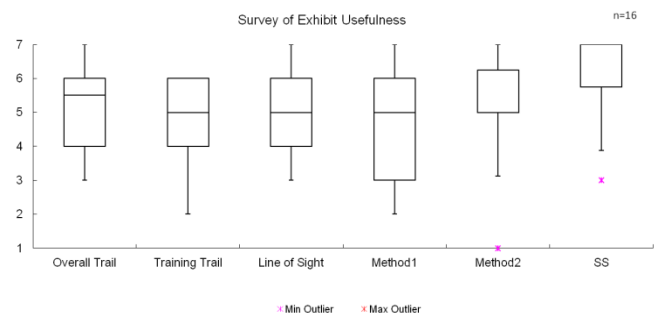


Figure 3. Exhibit usefulness results.

Presence Questionnaire. A presence questionnaire adapted from Witmer & Singer (1998) was given to the class in the experimental group, and the responses were condensed into one of four factors: control factors (how easy was it to control the virtual environment?), sensory factors (how immersive was the audio and video?), distraction factors (how aware were you of the external environment?), and realism factors (how often did you feel a part of the world?). Figure 4 shows the results of the questionnaire.

Students generally felt only somewhat immersed in the TMA trail. Reasons for this included insufficient workstations for each student (so two students had to share one

workstation) and some missing technologies that would have kept students immersed. The instructor often referenced objects in-world by pointing to the front projection screen, which caused all students to focus there. The development of an “in-world laser pointer” would allow for students to remain focused on their own monitor. When immersion and presence are broken down into the four questionnaire components, there is no major difference between how control, sensory, distraction, and realism factors contributed to the overall feeling of presence according to a 1x4 repeated-samples ANOVA ($F(3,56) = 0.46, p = .71$).

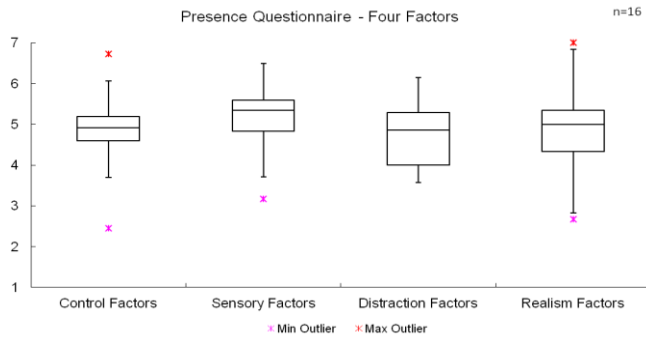


Figure 4. Presence questionnaire scores.

Test Scores. Pre, post, and retention test scores were collected across all groups. As noted previously, there were two Control-No Self Study (NSS) groups because one group had an invalid pre-test. A table of test scores (and standard errors) can be found in Table 1 and visualized in Figure 5.

	Pre-Test	Post-Test	Retention
Experimental	54.17% (3.15)	74.45% (3.21)	73.90% (2.37)
Self-Study	56.06% (1.72)	68.72% (3.08)	73.53% (3.93)
Control-NSS-1		69.12% (2.71)	70.36% (4.19)
Control-NSS-2	45.37% (2.02)	60.46% (4.02)	

Table 1. Test Scores across classes.

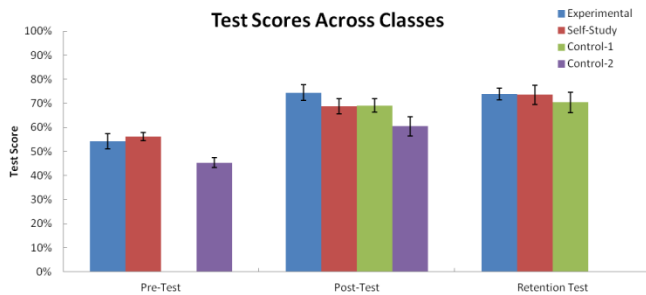


Figure 5. Graph of raw test scores across classes.

A learning gain score was calculated using the formula $(\text{Post-Pre}) / (1 - \text{Pre})$ in order to account for knowledge brought in at the earlier test and compensate for the restriction of variance in scores. This score was used in statistical analyses and visualized in Figure 6.

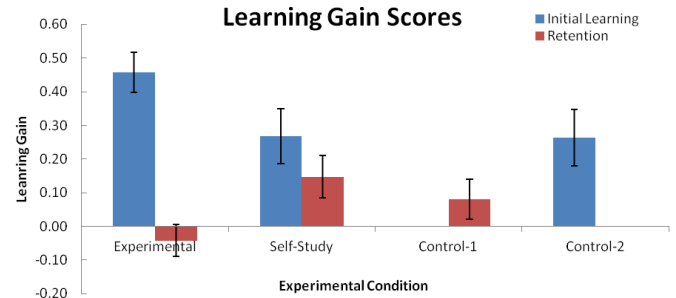


Figure 6. Learning gain scores across classes.

Any statistics performed on this data must be taken in the context that class sizes were unequal and also small (9-16 students). However, basic *t*-tests and descriptive statistics can still provide some insight into performance.

Both experimental and control classes gained some knowledge about TMA fundamentals during the approved, pre-existing two week module. Bonferroni-corrected *t*-tests ($p = 0.025$) demonstrate a trend in the level of initial learning ($t(10) = 2.08, p = 0.06$ for experimental vs. self-study groups, $t(8) = 2.12, p = 0.07$ for experimental vs. NSS groups) or retention ($t(10) = 1.51, p = 0.16$ and $t(8) = 0.97, p = 0.36$, respectively). There may be a small trend of students in the experimental group learning more initially thanks to the TMA trail. However, by the retention exam, the control groups had caught up, possibly due to additional study time and time spent in realistic combat system trainers.

It is possible that certain exhibits within the TMA trail were able to increase learning or retention more than others. In order to examine if this was the case, each test question was classified as pertaining to Line of Sight (LOS), Method 1 (Method1), Method 2 (Method2), or the Solution Sensitivity (SS). Average learning gain scores for each of these four areas were calculated for each student in each class, and then an average was taken across the students.

Figure 7 is a graph of initial learning gain for each of the TMA subareas. Students learned about LOS in previous class modules, so they did well on the LOS pre-test questions, which explains the lack of improvement. There was greater improvement on the Method 2 and SS questions, suggesting that the content improved learning of these concepts.

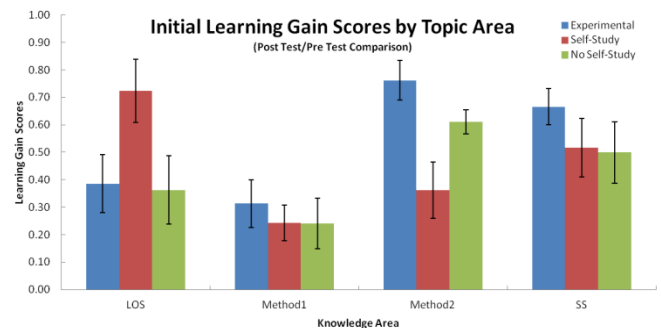


Figure 7. Initial learning scores separated by topic area.

T-tests (Bonferroni-corrected $p = 0.0125$) comparing the experimental group against both control groups demonstrated some statistical differences. For Method 2, the experimental

group gained more knowledge than the self-study group ($t(10) = 3.80, p < 0.005$), but scores were similar to the no-self-study group ($t(10) = 1.58, p = 0.15$). For the SS knowledge area, there was little evidence for a pattern ($t(10) = 1.28, p = 0.23$ for experimental vs. self-study, $t(10) = 1.43, p = 0.19$ for experimental vs. no-self-study), likely due to the large variance in learning gain scores.

Based on these (admittedly tenuous) results, it appears that four hours of immersion into virtual content helped students learn more about complex TMA concepts. The data suggests that creating immersive visualization for difficult-to-understand content can provide increased learning.

Individual Differences. One hypothesis is that the virtual content helped different students differently. Therefore, there may be a relationship between individual differences (e.g., ASVAB scores, age, video game experience) and improvement in scores between pre and post-test. A correlation was performed between student ASVAB scores and score improvement between pre and post-tests and post-and retention tests. Neither correlation was significant, $r(14) = 0.31, p = 0.25$; $r(14) = -0.30, p = 0.26$, indicating that ASVAB scores were unrelated to performance improvements. Additional analyses determined that participant age or hours playing video games (Green & Bavelier, 2003) were not correlated with performance (highest $r(14) = -0.30$, lowest $p = 0.26$). It is important to note that, with only sixteen students in the analysis, there may not have been enough participants to find an effect of individual differences.

DISCUSSION

The results from this TMA trail assessment must be put in context with the restrictions placed on the assessment – most notably, the limited number of students, the amount of time that the students were able to spend in the content, and the fact that students had to pair up and use workstations. Despite this, several valuable lessons were learned that will assist in improving immersive content creation and assessment.

Virtual content has a specific place in the curriculum. Since the TMA trail was designed to teach fundamental TMA concepts, the content can reinforce initial classroom learning and provide an introduction to displays seen in the physical trainers. In a previous unreported pilot study, students used the immersive content after using high-fidelity physical trainers, and the value of the content was not nearly as high. For the classes studied here, the TMA trail seemed to provide a positive stepping stone between lecture material and its application in the trainers. While scenario simulation may aid in the rehearsal and practice of knowledge, immersive exhibits can best be used to introduce and reinforce initial learning.

The primary advantages of virtual worlds are their emphasis on immersion and interactivity. Therefore, content must be built that allows students to be immersed and allows interaction. Students should be drawn into their own monitor, and there should be some kind of interactive portion of every exhibit designed to reinforce the concepts being taught. To this end, each student should have access to their own workstation, capable of running the virtual world client software reasonably well. Additionally, the ability to snap

students' cameras to the instructor's current view was useful in helping students tune out external distractions.

Finally, developing an "in-world laser pointer" would allow the instructor to reference objects while in the world, instead of pointing to the classroom projection screen. This technology would also enable remote instruction, where the instructors and students do not have to be collocated, but instead interact exclusively through the virtual world. In this scenario, students must be maximally immersed in the virtual world, as the external environment is almost wholly irrelevant.

Choosing what content to develop content for in the virtual world is important because, while content creation is relatively easy, the entire design, implementation, and integration process is still resource-consuming. While TMA concepts benefit from visualization, other aspects, such as the mathematics behind the algorithms, do not. The math is taught in the current curriculum at the schoolhouse via slides and text, and this is likely the best method for delivering the content. Results demonstrated the biggest value for time spent in the virtual world when complex concepts were involved.

Along these lines, it is important to identify discrete concepts and understand how they are connected for maximum efficacy. The paradigm of "exhibits along a trail," which is how the TMA trail is organized, is an implementation of knowledge scaffolding. Scaffolding is an idea where learning consists of building complex concepts on top of simpler ones (Yelland & Masters, 2007). A task analysis is a useful tool for breaking down complex tasks or concepts into its fundamental steps. This can be done through instructor interviews, documentation review, and observation.

While this assessment was limited in several ways, there was still a great deal of value in having students spend four hours in the virtual content. Students felt that the virtual world interface was easy to use, simple to navigate, and that the content varied between somewhat useful to very helpful. Despite having minimal time to interact with the virtual world, students were able to become comfortable inside the virtual content, understand the value of the capability, and also gain insight into some difficult-to-understand concepts.

REFERENCES

- Green, S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423, 534-537.
- Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels*. San Francisco, CA: Berrett-Koehler Publishers.
- Murphy, C. (2011). *Why Games Work – The Science of Learning*. Paper presented at the 2011 Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.
- Murphy, C., & Hussain, T. (2009). *Damage Control Trainer – A Serious Game Case Study*. Paper presented at the 2009 United States Coast Guard Human Performance Technology Workshop, Williamsburg, VA.
- Witmer, B., & Singer, M. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence*, 7, 225-240.
- Yelland, N., & Masters, J. (2007). Rethinking scaffolding in the information age. *Computers and Education*, 48, 362-382.